Hands-on-ML with R

Koji Mizumura

October 29th,2018 -

Table of Contents

Machine learning continues to grow in importance for many organizations across nearly all domains. Examples include:

* predicting the likelihood of a patient returning to the hospital (*readmission*) within 30 days of discharge,
* segmenting customers based on common attributes or purchasing behavior for target marketing,
* predicting coupon redemption rates for a given marketing campaign,
* predicting customer churn so an organization can perform preventative intervention,
* and many more!

In essence, these tasks all seek to learn from data. To address each scenario, we use a given set of *features* to train an algorithm and extract insights. These algorithms, or *learners*, can be classified according to the amount and type of supervision provided during training. The two main groups this book focuses on includes: ***supervised learners*** that are used to construct predictive models, and ***unsupervised learners*** that are used to build descriptive models. Which type you will need to use depends on the learning task you hope to accomplish.

## Supervised Learning

A ***predictive model*** is used for tasks that involve the prediction of a given output using other variables and their values (*features*) in the data set. Or as stated by @apm, predictive modeling is *“the process of developing a mathematical tool or model that generates an accurate prediction”* (p. 2). The learning algorithm in a predictive model attempts to discover and model the relationship among the ***target*** response (the variable being predicted) and the other features (aka predictor variables). Examples of predictive modeling include:

* using customer attributes to predict the probability of the customer churning in the next 6 weeks,
* using home attributes to predict the sales price,
* using employee attributes to predict the likelihood of attrition,
* using patient attributes and symptoms to predict the risk of readmission,
* using production attributes to predict time to market.

Each of these examples have a defined learning task. They each intend to use attributes () to predict an outcome measurement ()

Throughout this text I will use various terms interchangeably for:

* : “predictor variables”, “independent variables”, “attributes”, “features”, “predictors”
* : “target variable”, “dependent variable”, “response”, “outcome measurement”